LEVEL II

Research Memorandum 70-4

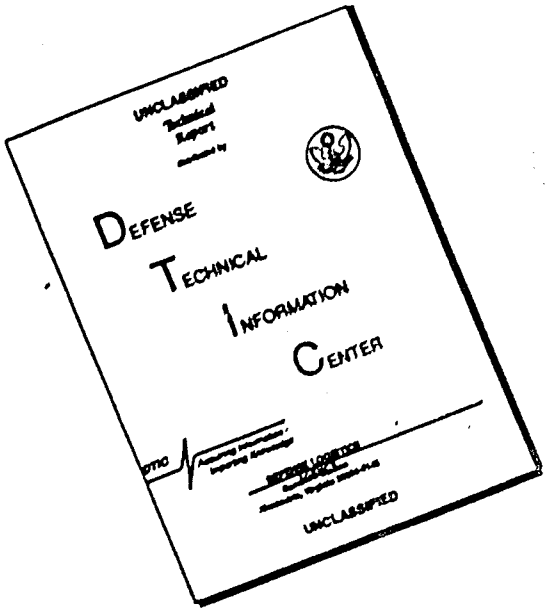
INTERACTION BETWEEN PSYCHOLOGICAL TESTS AND CURRICULA IN THE MARINE HULL REPAIRMAN COURSE

BESRL

U. S. Army
Behavior and Systems Research Laboratory

August 1970

DISCLAIMER NOTICE



THIS DOCUMENT IS BEST QUALITY AVAILABLE. THE COPY FURNISHED TO DTIC CONTAINED A SIGNIFICANT NUMBER OF PAGES WHICH DO NOT REPRODUCE LEGIBLY.

Army Project Number
16 20[illegible]A[illegible]

Differential Classification a-11

7 Research Memo[illegible]m 70-4

6 INTERACTION BETWEEN PSYCHOLOGICAL TESTS AND CURRICULA IN THE MARINE HULL REPAIRMAN COURSE

10 Roger L. Williamson

Milton H. Maier, Task Leader

14 BESRL-R.M.-70-4

12 11

Submitted by:
Edmund F. Fuchs, Chief
Military Selection
Research Division

Approved by:
J. E. Uhlaner, Director
Behavior and Systems
Research Laboratory

11 Aug[illegible]70

INTERACTION BETWEEN PSYCHOLOGICAL TESTS AND CURRICULA IN THE MARINE HULL REPAIRMAN COURSE

BACKGROUND

Several Army training courses have been revised by changing the training methods and, to a lesser extent, the course content. The method of training in the revised courses places emphasis on learning by doing rather than on the lecture-demonstration-performance of traditional courses. Also, content of the revised courses emphasizes performance of specific tasks rather than comprehension of general principles. One consequence may be a less flexible graduate who needs retraining as equipment is modified.

The results of recent research on Army courses showed that general aptitude, as measured by tests of arithmetic reasoning and verbal ability, has high validity in predicting success in virtually all types of Army training. The presence of this general predictor suggested that the traditional course placed a high demand on the ability to work with concepts. With the reduced level of abstraction in the revised courses, there may be less need for general learning ability, and more need for specific job-related abilities that could reflect a different pattern of aptitudes for success.

An ongoing study conducted by BESRL's Differential Classification Work Unit is aimed at gathering evidence about differences in validity patterns of Army classification tests and some experimental tests for courses of the two types--traditional and revised. More specifically, what are the nature and magnitude of interaction effects between training course revisions and the validity of tests used for predicting success in the Marine Hull Repairmen Course (44K20)?

METHOD

The criterion measures were appropriate for the particular training methods and content used in each curriculum. In the control sample, which received the traditional academically oriented course, the final course grade was heavily weighted by a multiple choice pencil-and-paper test. In the experimental sample, which received a more performance oriented course, the final course grade was composed of three parts. The first part was a general performance score based on a series of evaluations used for instruction purposes as well as for testing. While the man worked on a given project, the instructor corrected his mistakes and pointed out areas in which he was progressing satisfactorily. The quality of the project and the performance of the man were graded at the termination of an allotted time. The second part of the grade was a score on specific performance based on a series of tests of the man's ability to perform complicated tasks to rigid specifications, employing the methods used in a major area of the course such as marine pipefitting.

No instruction was given during the test. Each step of the task was observed by an instructor and graded as it was performed. The test was terminated and the project was graded at the end of a given period of time. The last part of the final course grade was a written objective test.

SAMPLES AND PROCEDURES

The U. S. Army Transportation School, Fort Eustis, Virginia, provided data on samples of men who had taken the Marine Hull Repairman course under the two curricula. A control sample consisted of 139 men enrolled in the Marine Hull Repairman Course in 1967 under the traditional academically oriented training curriculum. The experimental group was made up of 56 men enrolled in the same course in 1968 under the revised performance-oriented training curriculum. Since some of the training methods used in the revised performance-oriented curriculum (experimental sample) were introduced into the course on a gradual basis, differences between the control and experimental samples were somewhat reduced.

Data on the control sample were available for only a limited number of predictors and one criterion measure. The experimental sample had data available not only for the Army Classification Battery (ACB) tests but also for a number of other tests not usually given to Army personnel. Table 1 displays the variables for which data were available for each sample. Data for five ACB predictors (the Verbal, Arithmetic Reasoning, and Pattern Analysis tests and the Electronics and Motor Maintenance Aptitude Areas) and one criterion (final course grade) were available for both samples. The table also shows the non-operational variables included in the analysis.

A number of statistical analyses were performed. A regression analysis of the Arithmetic Reasoning (AR) test predicting the final course grade was computed for the control and experimental samples separately. The purpose of this analysis was to compare the regression of final course grade on AR in the two training methods, since the test has had consistently high validity over the years. The second statistical procedure was to calculate an intercorrelation matrix for each sample of all available predictor and criterion variables. A multiple regression equation was also calculated for each sample separately, using only those predictors available for both samples against the final course grade.

Table 1.

PREDICTOR AND CRITERION DATA AVAILABLE FOR SAMPLES[a]

Data	Experimental Revised Course	Control Traditional Course
Predictors		
Education	X	
ACB Tests		
Verbal (VE)	X	X
Arithmetic Reasoning (AR)	X	X
Shop Mechanics (SM)	X	
Pattern Analysis (PA)	X	X
Army Clerical Speed Aptitude (ACS)	X	
Automotive Information (AI)	X	
Mechanical Aptitude (MA)	X	
Electrical Information (ELI)	X	
Classification Inventory (CI)	X	
Army Radio Code Aptitude (ARC)	X	
Electronics Aptitude Area [(MA+2ELI)/3]	X	X
Motor Maintenance AA [(MA+2AI)/3]	X	X
Wechsler Adult Intelligence Scale (WAIS) (14 scores)	X	
Iowa Silent Reading Test (1 score)	X	
Prognostic Test of Mechanical Ability (1 score)	X	
Domino Test (1 score)	X	
Survey of Interpersonal Values (Gordon) (6 scores)	X	
Criteria		
Course Grades:		
General Performance	X	
Specific Performance	X	
Objective Examination	X	
Final	X	X

[a] "X" indicates data available.

RESULTS AND DISCUSSION

In the experimental sample, the specific performance score had extremely high correlation with final course grade (r = .92, Table 2); the general performance score and the objective test had lower coefficients (.74 and .42, respectively). It would appear, then, that a man's standing in his class was fairly well defined by his ability to perform the specific tasks. Conversely, the objective test carried only the weight due to the part-whole correlation reflecting its presence in the final course grade, desp te its high standard deviation.

Table 2

MEANS, STANDARD DEVIATIONS, AND INTERCORRELATIONS OF CRITERION SCORES FOR THE EXPERIMENTAL SAMPLE (N = 56)

Criterion	Mean	S.D.	Intercorrelations			
Final Course Grade	85	4	FCG			
General Performance	86	4	.74	GP		
Specific Performance	81	6	.92	.54	SP	
Objective Test	80	12	.42	.38	.26	OBJ

Table 3 shows all the valitidy coefficients obtained in both samples. In the experimental sample, the cognitive predictors attained highest validity with the general performance rating and lowest validity with the specific performance ratings. Coefficients are shown for measures outside the ACB and for education simply as a matter of record, inasmuch as the data were analyzed.

The highest validity coefficient against the specific performance score (r = .43) and against final course grade (r = .40) in the experimental sample was attained by the only noncognitive test in the ACB, the Classification Inventory (CI). The CI is a self-description inventory keyed to identify men who are well-adjusted and cooperative in their attitudes toward rigorous Army training and Army life. Individuals in the experimental sample who had a high CI score may also have received a high performance score because of their close cooperation with the instructor and performance rater. Two scales in Gordon's Survey of Interpersonal Values, Benevolence and Leadership, indicated some support for the CI validity (r = .40). The Benevolence Scale, which measures traits opposite to those of the CI (r = -.21) and the Leadership Scale (r = -.47) had a moderate negative correlation with final course grade (r = -.37). The Leadership Scale had moderate positive correlation with the CI (r = .29) and with final course grade (r = .36).

Table

VALIDITY COEFFICIENTS FOR EXPERIMENTAL AND CONTROL SAMPLES[a]

	Criteria				
	Final Course Grade		Gen Perf Score	Specific Perf Score	Obj Test
Predictors	Exp	Con	Exp	Exp	Exp
ACB					
Verbal	4[illegible]	44	[illegible]	[illegible]	37
Arithmetic Reasoning	40	[illegible]	[illegible]	30	35
Shop Mechanics	40		[illegible]7	25	40
Pattern Analysis	37	37	62	23	33
Army Clerical Speed Aptitude	31		30	25	23
Automotive Information	39		54	30	33
Mechanical Aptitude	43		[illegible]8	26	42
Electronic Information	34		43	23	44
General Information	4[illegible]		[illegible]	34	63
Classification Inventory	4[illegible]		51	43	19
Army Radio Code Aptitude	28		40	17	41
Electronics Aptitude Area	3[illegible]	61	[illegible]0	25	49
Motor Maintenance Aptitude Area	43	5[illegible]	5[illegible]	30	40
WAIS					
Verbal	47		5[illegible]	3[illegible]	53
Information	31		35	2[illegible]	52
Comprehension	43		44	37	35
Arithmetic	31		45	20	55
Similarities	4[illegible]		5[illegible]	33	39
Digit Span	32		3[illegible]	26	2[illegible]
Vocabulary	35		4[illegible]	26	51
Performance	36		61	23	38
Digit Symbol	14		38	-03	1[illegible]
Picture Completion	33		56	26	30
Block Design	43		60	32	36
Picture Arrangement	29		42	23	21
Object Assembly	12		28	0[illegible]	2[illegible]
Full	44		63	33	50
MISC.					
Iowa Silent Reading	38		37	2[illegible]	55
Mechanical Abilities	42		67	26	56
Domino	3[illegible]		63	2[illegible]	38
Education	25		3[illegible]	12	21
Gordon SIV					
Support	-0[illegible]		-03	-15	14
Conformity	-15		-2[illegible]	-05	-24
Recognition	00		-07	01	03
Independence	10		25	-03	23
Benevolence	-37		-40	-20	-42
Leadership	36		24	30	16

[a] Decimal points omitted.

A comparison of the control and experimental samples with respect to the regression of final course grade on AR appears in Figure 1. The mean final course grade in each AR interval is indicated in the AR column for each sample. Trend lines in both samples showed a clear linear relationship with nearly identical slopes, indicating that the AR test functioned in both samples equally well as a predictor throughout the score range. Differences between the trend lines indicated either better performance or relaxed standards in the experimental course.

Zero-order correlation and multiple regression analyses were performed using five predictors against the final course grade for each sample. The predictors were the Verbal (VE), Arithmetic Reasoning (AR), and Pattern Analysis (PA) tests, plus the Electronics (EL) and Motor Maintenance (MM) Aptitude Area composites. Data for these variables are summarized in Table 4.

Within the experimental sample, the zero-order validity coefficients for the five predictors shown in Table 4 were of the same general magnitude, with MM the highest (r = .43) and PA the lowest (r = .37). The beta weights in the experimental sample also indicated a stronger contribution of MM to the prediction of final course grade, as would be expected in a course emphasizing more practical mechanics. The validity coefficients were uniformly lower in the experimental sample than in the control sample, perhaps as a result of differences in the criterion measures. Within the control sample, the validity coefficients were more varied. MM (r = .59) and EL (r = .61) were the best of the five predictors even though the curriculum for the control sample was more academically oriented. PA, again, had the lowest of the five (r = .37). The beta weights for EL and MM in the control sample also pointed to a higher contribution of the mechanical predictors.

All five predictors showed significant validity in both samples. In terms of beta weights, the MM aptitude area and the AR test showed high independent contributions in both samples. VE was also significant in the experimental sample and EL carried the highest weight in the control sample. The multiple R of .48 in the experimental sample was not significally different at the .05 level from the multiple R of .66 in the control sample. The relatively high negative beta weight of -.12 for VE in the control sample may be peculiar to the sample. The data, then, did not indicate that the reduced level of abstraction in the revised course required a markedly different pattern of aptitudes for success.

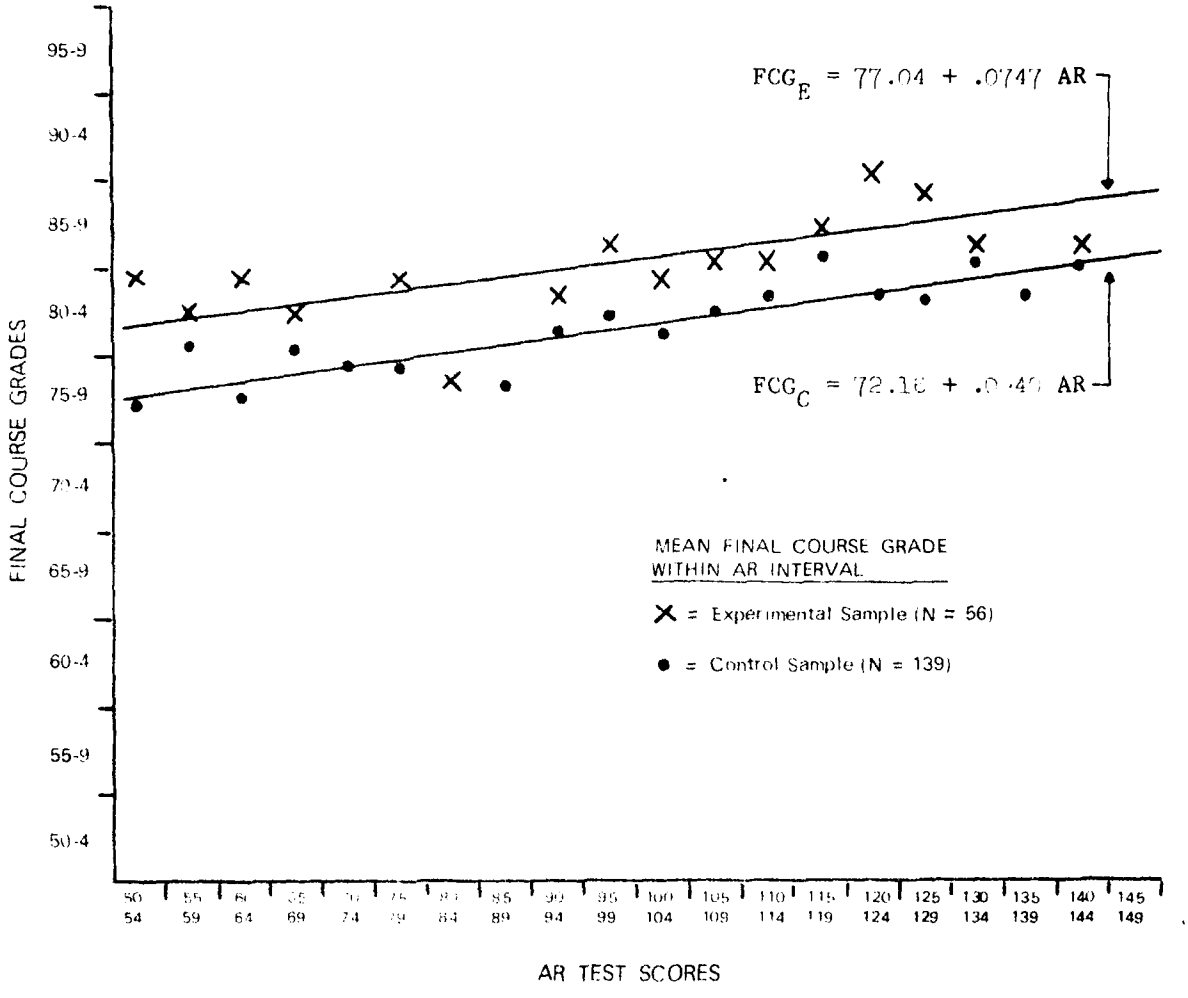


Figure 1. Regression of final course grade on Arithmetic Reasoning test scores in the Marine Hull Repairman Course

Table 4

COMPARISON OF EXPERIMENTAL AND CONTROL SAMPLES
ON FIVE PREDICTORS AND ONE CRITERION

Predictors	Mean		S.D.		Validity Coefficients		Multiple R		Beta Weight	
	Exp[a]	Con[a]	Exp	Con	Exp	Con	Exp	Con	Exp	Con
VE	100	97	24	24	.40	.44	.47	.657	.14	-.12
AR	100	97	22	20	.40	.55			.15	.24
PA	105	106	24	19	.37	.37			.01	-.02
EL[b]	104	103	20	20	.39	.61			.01	.35
MM[b]	108	105	19	19	.43	.57			.25	.25
Criterion										
FCG	85	81	4.0	3.5						

[a] Experimental sample size = 56; Control sample size = 139.
[b] EL Aptitude Area score = (MA + 2ELI)/3; MM Aptitude Area score = (MA + 2AI)/3.

SUMMARY

The training methods and course content in the Marine Hull Repairman Course have been revised by placing greater emphasis on learning through actual performance of specific tasks than on learning through presentation of the more abstract principles linked to the tasks. In order to obtain an indication of the nature and magnitude of the interaction effect between training methods and tests, the validity patterns of tests for a sample of men (N = 139) who took the previously standard acdemically oriented course were compared with the validity patterns of tests for a sample of men (N = 56) who took the revised performance oriented course. While for the revised course sample the full ACB and a number of other test scores were available as predictors, the standard course sample had only five scores whose validities could be compared across the samples.

Zero-order correlation and multiple regression analyses of the five common predictors against the final course grades did not uncover significant differences in validity patterns between the two samples to support the hypothesis of greater relative validity in the control sample for the academic predictor tests (VE and AR), or greater relative validity in the experimental sample for the mechanically oriented aptitude areas (EL and MM). The highest validity coefficient against the specific performance

rating and the final course grade in the experimental sample was not one obtained through a test of mechanical abilities but was attained by the only noncognitive test in the ACB, the Classification Inventory (CI). If the high validity for CI proves to be stable in subsequent studies, a new dimension in the criterion for the revised course will have been indicated.